



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/757,313	01/14/2004	Xiang Tong	940630-010-024	3516
<div>7590      06/12/2007</div> <div>Blaney Harper 51 Louisiana Avenue, N.W. Washington, DC 20001-2113</div>			<div>EXAMINER</div> <div>NEWAY, SAMUEL G</div>	
			<div>ART UNIT</div> <div>2626</div>	<div>PAPER NUMBER</div>
			<div>MAIL DATE</div> <div>06/12/2007</div>	<div>DELIVERY MODE</div> <div>PAPER</div>

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<p align="center"><b>Office Action Summary</b></p>	<p><b>Application No.</b></p> <p align="center">10/757,313</p>	<p><b>Applicant(s)</b></p> <p align="center">TONG ET AL.</p>	
	<p><b>Examiner</b></p> <p align="center">Samuel G. Neway</p>	<p><b>Art Unit</b></p> <p align="center">2626</p>	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 14 January 2004.
- 2a) ☐ This action is **FINAL**.                      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-16 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-16 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 14 January 2004 is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)          | 4) <input type="checkbox"/> Interview Summary (PTO-413)           |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date: _____                                      |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)          | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date: _____   | 6) <input type="checkbox"/> Other: _____                          |

### **DETAILED ACTION**

1. This is responsive to the Application filed on 14 January 2004.

#### ***Drawings***

2. The drawings are objected to because in FIG. 2, it is believed there should be two arrows pointing from item 216, one representing the affirmative while the other representing the negative to the question "Is N-gram in Language Database".

Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of an amended drawing should not be labeled as "amended." If a drawing figure is to be canceled, the appropriate figure must be removed from the replacement sheet, and where necessary, the remaining figures must be renumbered and appropriate changes made to the brief description of the several views of the drawings for consistency. Additional replacement sheets may be necessary to show the renumbering of the remaining figures. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

***Claim Objections***

3. Claim 2 is objected to under 37 CFR 1.75(c), as being of improper dependent form for failing to further limit the subject matter of a previous claim. Applicant is required to cancel the claim(s), or amend the claim(s) to place the claim(s) in proper dependent form, or rewrite the claim(s) in independent form. Each limitation in claim 2 is already disclosed in claim 1 on which claim 2 depends.
4. Claim 9 is objected to because it is believed "method of determining" at the beginning of the preamble should be 'A method of determining'.

***Claim Rejections - 35 USC § 112***

5. The following is a quotation of the second paragraph of 35 U.S.C. 112:  

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.
6. Claims 1 – 6 and 9 – 11 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 1 recites the limitation "said n-grams" in (c). It is unclear as to which n-gram, the textual passage n-grams or the databases n-grams, the limitation refers. The Examiner will read the limitation as referring to the textual passage n-grams. Claims 2 – 6 also recite the limitation "said n-gram" and raise the same issue as claim 1 on which they all depend.

Claim 9 recites the limitation "said short words" in (c). It is unclear as to which short words, the textual passage short words or the databases short words, the limitation refers. The Examiner will read the limitation as referring to the textual passage short words.

Claim 10 recites the limitations "said n-grams and said short words" in (c). It is unclear as to which n-gram and said short words, the textual passage n-grams and said short words or the databases n-grams and said short words, the limitations refer. The Examiner will read the limitations as referring to the textual passage n-grams and said short words.

Claim 11 recites the limitation "said n-grams" in (d). It is unclear as to which n-gram, the textual passage n-grams or the databases n-grams, the limitations refers. The Examiner will read the limitation as referring to the textual passage n-grams.

### ***Claim Rejections - 35 USC § 103***

7. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

8. Claims 1 – 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Schulze (USPN 6,167,369) in view of Messerly et al (USPN 6,076,051).

Claim 1:

Schulze discloses a method of determining the language of a textual passage, the method comprising the steps of:

(a) parsing said textual passage into a plurality of n-grams ("tokenizes large samples of text ", col. 1, lines 18-34);

(b) comparing each of said n-grams with a plurality of databases, wherein each of said databases comprises a list of n-grams associated with a specific language ("The probabilities are then used to guess the language of a sentence", col. 1, lines 18-34);

(c) determining an initial weight for each of said n-grams, per language, by calculating the frequency with which each of said n-grams appears in each of said databases and dividing said frequency by the total number of n-grams in said respective database ("probability of a retained trigram is approximated by summing the frequency of all retained trigrams for the language and dividing the trigram's frequency by the sum of frequencies", col. 1, lines 18-34).

However, Schulze does not explicitly disclose determining the number of databases containing each n-gram and dividing the n-grams initial weight by said number.

In a similar information retrieval method, Messerly discloses

(d) determining the number of said databases within which each of said n-grams (token combination) appear ("inverse document frequency", col. 12, equation 1);

(e) altering said initial weight for each of said n-grams by multiplying said initial weight with the inverse of said number of databases within which each of said n-grams

Art Unit: 2626

appear ("The facility preferably uses a combination of inverse document frequency ... to rank the matching target documents", col. 12, lines 36-39).

It would have been obvious to one with ordinary skill in the art at the time of the invention to alter the n-grams initial weight in Schulze's method using inverse document frequency because that would give "greater weight to a token combination appearing in fewer of the targets documents" (Messerly, col. 12, lines 40-43).

Schulze further discloses

(f) producing the weight of each language over the text passage by calculating, per language, the sum over each n-gram in the text passage of the products of the number of times that that n-gram appears in the text passage and the language-specific altered weight calculated in step (e) for that n-gram ("dividing the sentence into trigrams and calculating the probability of the sequence of trigrams for each language", col. 1, lines 18-34);

(g) sorting the list of per language passage weights from step (f) in decreasing order, returning the most likely language for the text passage as the first element (highest weight) in the list ("The language with the highest probability for the sequence of trigrams is chosen", col. 1, lines 18-34).

Claim 2:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein the step of determining an initial weight for each of said n-grams, per language, comprises the steps of calculating the frequency with which each of said n-grams appears in each of said databases and dividing said frequency by the total number of n-

Art Unit: 2626

grams in said respective database ("probability of a retained trigram is approximated by summing the frequency of all retained trigrams for the language and dividing the trigram's frequency by the sum of frequencies", col. 1, lines 18-34).

Claim 3:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said n-grams have a size selected from the group consisting of bi-grams, tri-grams, and quad-grams (col. 16, lines 16-21).

Claim 4:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said n-grams are anchored n-grams (col. 16, lines 16-21).

Claim 5:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said n-grams are replacement-type n-grams (col. 16, lines 16-21).

Claim 6:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said n-grams are any combination of n-grams, including anchored n-grams and/or replacement-type n-grams, and/or n-grams of different lengths (col. 16, lines 16-21).

Claim 7:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said textual passage comprises 20 or more words (col. 6, lines 62-65).

Claim 8:



Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein said textual passage comprises 40 or more words (col. 6, lines 62-65).

Claim 9:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein the n-grams are words (col. 1, lines 34-46).

Claim 10:

Schulze and Messerly disclose the method of claim 1, Schulze further discloses wherein the n-grams are a combination of n-grams and words (col. 2, lines 55-64).

Claim 11:

Claim 11 is similar in scope and content to claim 1 and is rejected with the same rationale.

Claim 12:

Schulze and Messerly disclose the system of claim 11, Schulze further discloses a scanner and an optical character recognition device, wherein said scanner and said optical character recognition device are connected to said central processing unit, wherein said program receives a textual passage from a document scanned by said scanner (col. 17, line 65 to col. 18, line 10).

Claim 13:

Schulze and Messerly disclose the system of claim 11, Schulze further discloses wherein said program comprises a user interface that allows a user to enter said textual passage (col. 18, lines 11-12).

Claim 14:

Schulze and Messerly disclose the system of claim 13, Schulze further discloses wherein said user interface is a graphical user interface (col. 18, lines 11-15).

Claim 15:

Schulze and Messerly disclose the system of claim 13, Schulze further discloses wherein said user interface displays the identified language (col. 18, lines 11-15).

Claim 16:

Schulze and Messerly disclose the system of claim 11, Schulze further discloses wherein said program comprises a user interface that allows a user to enter a Uniform Resource Locator that contains said textual passage (col. 17, lines 51-64).

***Conclusion***

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

- a. Damashek (USPN 5,418,951) discloses a method of identifying, retrieving, or sorting documents by language or topic involving the steps of creating an n-gram array for each document in a database, and parsing an unidentified document or query into n-grams.
- b. de Campos (USPN 6,272,456) discloses method for identifying a language of a document from a small sample input of the document by using n-gram profiles.
- c. Cavnar et al ("N-Gram Based Text Categorization", Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information

Art Unit: 2626

Retrieval, 1994) discloses a method for identifying the language of a text using n-gram based text categorization.

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Samuel G. Neway whose telephone number is 571-270-1058. The examiner can normally be reached on Monday - Friday 8:30AM - 5:30PM EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R Hudspeth can be reached on 571-272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

SN

SN

  
DAVID HUDSPETH  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2626